HIDE AND SEEK: Steganography in Digital Forensics

What is steganography?

The shortest definition for steganography is *to conceal data or messages inside non secret messages or data*. Which makes it a form or offshoot of cryptography. However, cryptography changes the information to ciphertext which cannot be understood without a decryption key. So, if someone were to intercept this encrypted message, they could easily see that some form of encryption had been applied. On the other hand, steganography does not change the format of the information but it conceals the existence of the message.

In one modern form, steganography is a methodology of hiding information in the unnecessary pixels of a picture. While it is not terribly common, hackers have used it in a variety of ways, from hiding malware to sending commands and information and exfiltrating data. Steganography can be very difficult to detect as the image itself looks the same as the original. This makes steganography a very effective tool for phishing emails as a way to spread malicious files rather than attaching them as a file. Whether steganography is being used in phishing or as a way for malware to exfiltrate data, detecting steganography can be very difficult for a security operations center. However, building such a capability can greatly improve your security capabilities.

Origins of Steganography:

Steganography is the combination of two Greek words, *steganos* which means "covered, concealed, or protected", and *graphein* which means "writing". It has been said that this method was first followed by an ancient Greek king who would shave his most trusted servant's head, and tattoo a secret message there. After the hair had regrown, he would send that servant to convey the message that had information about an upcoming attack (http://csis.pace.edu/~ctappert/srd2005/d1.pdf).

Some say that the first recorded use of the term “steganography” was in 1499 by Johannes Trithemius in his Steganographia, a treatise on cryptography and steganography, disguised as a book about magic. Use of this practice continued through time, particularly during periods of war or revolution. For instance, it was commonly practiced by both the British and the Americans during the Revolutionary War with the use of invisible ink. It is known that during both world wars, female spies used knitting to send messages, perhaps making an

irregular stitch or leaving an intentional hole in the fabric. With morse code being a common transport of messaging in times of war, steganography was used as well as an additional form of protection. Messages would be written in morse code on yarn used to make clothing. Morse code could also be hidden underneath stamps on mailed correspondence. Also in the 20th century, the German invention of the microdot (photographs the size of a printed period allowing transmission of large amounts of data) was a breakthrough invention. (https://www.giac.org/paper/gsec/2474/information-hiding-art-steganography/104295)

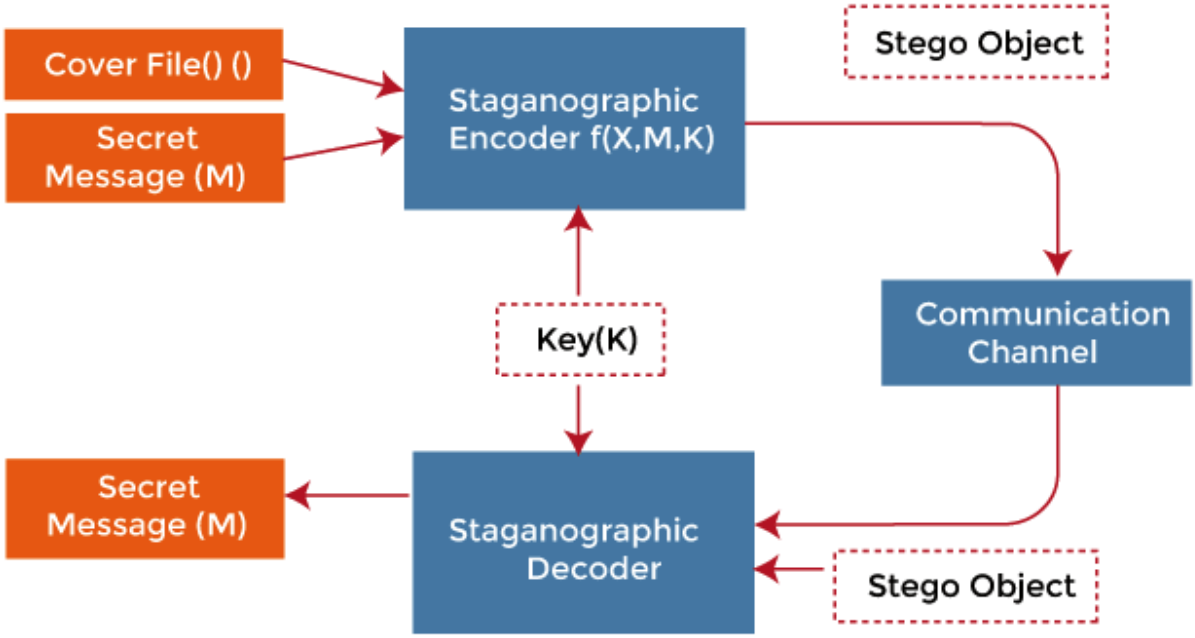
In more modern times the invention of computers has provided new and more sophisticated ways to hide and transport information. Since sometime in the 1980's personal computers began to be used for traditional steganography applications. While this particular development has been slow, a large number of steganography software exists today. Being a form of security through secrecy, the steganography algorithm, unlike a cryptographic algorithm, must consider the plausible form that the generated data must have, so that they do not cause suspicion. As information went digital, steganography changed. Messages could be hidden in the 1s and 0s of electronic files -- pictures, audio, video, executables, whatever. The hidden communications could even be slowly dribbled into the torrent of IP traffic. Compression schemes -- like JPEG for images or MP3 for audio -- introduce errors into the files, making a message even easier to hide. New colors or tones can be subtly added or removed, to cover up for the changes. (https://www.wired.com/2010/06/alleged-spies-hid-secret-messages-on-public-websites/)

It's been reported both before and after 9/11 that Al-Qaeda hid messages in digital pornography. In recent years law enforcement has tracked down spies for both China, Russia and more for using steganography to transmit stolen secret information. Today steganography therefore presents itself as an ideal tool for the creation of secret communication channels, which can be used in sophisticated scenarios of espionage, computer crime and violation of privacy of both public and private subjects.

How Files are created and hidden:

In digital steganography, electronic communications can include steganographic encoding within a transport layer, such as a document file, an image file, a program or a protocol. Multimedia files are ideal for steganographic transmission because of their large size. For example, a sender might send a harmless image file and adjust the color of one pixel in a hundred to match an alphabetic character. The change is so subtle that someone is unlikely to notice it unless they are specifically looking for it.

Figure 1.



https://www.javatpoint.com/image-steganography-using-python

Figure 2.

Image 1 (original)

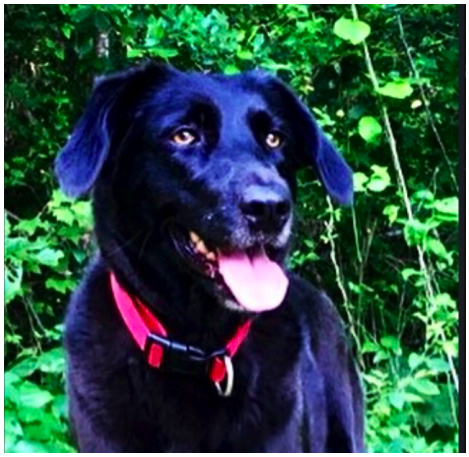


Image 2 (steg image)

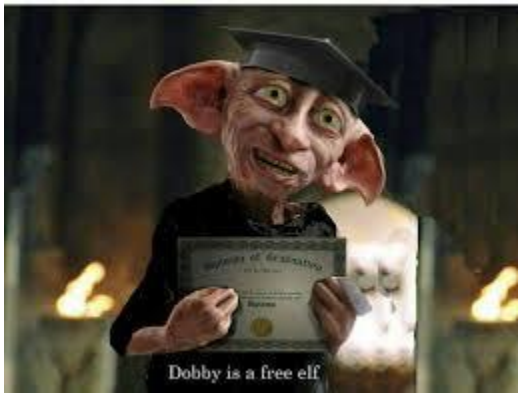
There are generally 5 main types of steganography techniques:

1. Text Steganography - hiding information inside text files. This can be changing the format of existing text, changing words within a text, generating random character sequences or using context-free grammars to generate readable texts. Various techniques used to hide the data in the text are: linguistically, statistical generation, and format based method.
2. Image Steganography - hiding data behind or within a cover object that is an image file. This is commonly done because there are a huge number of bits present in the digital representation of an image. There are a lot of ways to hide information inside an image. Common approaches to hide information inside images include: coding and cosine transformation, encrypt and scatter, redundant pattern encoding, least significant bit insertion, and masking and filtering.
3. Video Steganography - data can be hidden in digital video format. The advantage of this type is a large amount of data can be hidden inside and the fact that it is a moving stream of images and sounds. The main types of video steganography embedding data in uncompressed raw video and compressing it later and embedding data directly into the compressed data stream.
4. Audio Steganography - the secret message is embedded into an audio signal which alters the binary sequence of the corresponding audio file. Data can be hidden in WAV, AU, MP3 files. Hiding secret messages in sound is a much more difficult process when compared to others, like images. These methods include: phase coding, parity encoding, least significant bit encoding, spread spectrum.
5. Network Steganography (or protocol steganography) - this is the technique of embedding information within network control protocols used in data transmission such TCP, UDP, ICMP etc. For example, you can hide information in the header of a TCP/IP packet in some fields that are either optional.
   (https://www.edureka.co/blog/steganography-tutorial)

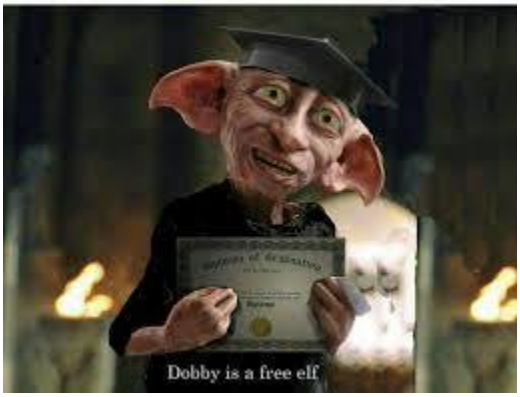
An Image Steganography Example:

The image on the left is an original and the image on the left has a hidden file embedded.

Figure 3.



Original



Steg

We took the original image and used the program Steghide to embed a hidden “secret file” within it.

Figure 4.

```
┌──(kali㉿kali)-[~/Documents]
└─$ steghide extract -sf dobbyog.jpeg
Enter passphrase:
the file "secretfile.txt" does already exist. overwrite ? (y/n) y
wrote extracted data to "secretfile.txt".

┌──(kali㉿kali)-[~/Documents]
└─$ cat secretfile.txt
```

We could then generate image comparators to see the difference in what look like identical images.

Figure 5.

Original

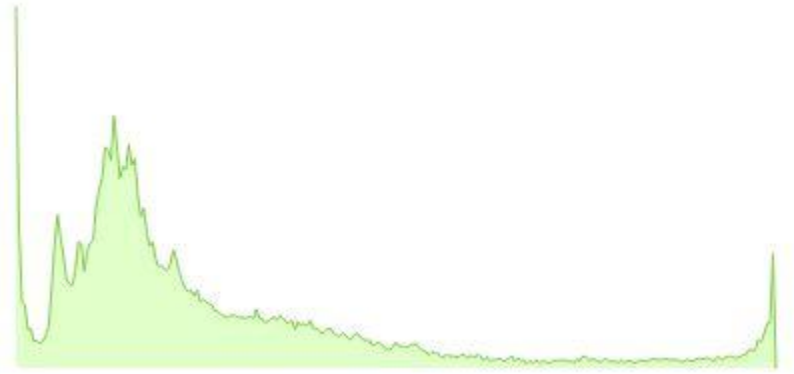
Filename:	dobby2.jpeg
Filetime:	2022-02-17 21:19:42 GMT
File Type:	image/jpeg
Dimensions:	300x241
Color Channels:	3
Unique Colors:	29,836
File Size:	11,444 bytes
MD5:	d90cae57f1e47ac558301dd325e3bbfb
SHA1:	76fcb9f60c33f3bdf35b361afeb6957840b74a77
SHA256:	83b18c4875e90e4d4783e0a721afec2ae26d7c1452fdd0eb26d3800b2fd16d1
First Analyzed:	2022-02-18 18:58:21 GMT

Steg

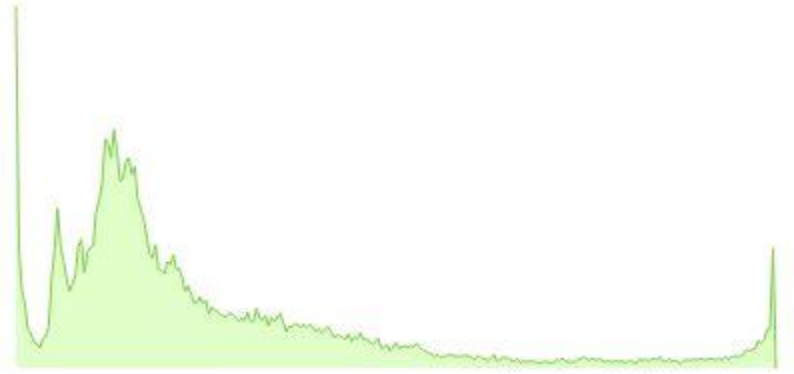
Filename:	dobby2.jpeg
Filetime:	2022-02-18 19:01:06 GMT
File Type:	image/jpeg
Dimensions:	300x241
Color Channels:	3
Unique Colors:	30,577
File Size:	11,483 bytes
MD5:	66e5adfeca479abb41e47edf34417635
SHA1:	629018b08f31bd705b3ce4267bae12232b7fff4f
SHA256:	dc269b0867621d078fbed4c47ba8c979613582cb28e6c496daff277ebbd3be65
First Analyzed:	2022-02-18 19:04:44 GMT

Figure 6.

Original



Steg



These histograms display the differences in the files. Slight, but noticeable.

Figure 7.

The pink dots shows the distribution of the data throughout the file.

Tools and Techniques to detect steg files:

Below are some of the main approaches to detecting data concealed in digital images:

1. Visual detection: Analyzing repetitive patterns may reveal the identification of a steganography tool or hidden information. To inspect these patterns an approach is to compare the original cover image with the stego image and note visible differences. This is called a known-carrier attack.

2. Signature-based analysis: identifiable information left inside files after data has been hidden in the file. These signatures are unique to a particular steganography program and help in steganalysis. This is also useful when the original image is not available.

3. Statistical analysis: the method of detecting steganography by identifying or observing data that differs significantly from the majority of the data being analyzed.

4. Structural analysis: to investigate for data hidden inside images by performing a simple analysis of the carrier file when we have both the original and the suspected stego files.

(https://www.sciencedirect.com/topics/computer-science/steganography-tool)
(https://www.wetstonetech.com/service/steganography-analysis-and-forensics/)

Per our example in previous sections the secret file for our project was not too difficult to retrieve using the same steganography tool.

Figure 8.

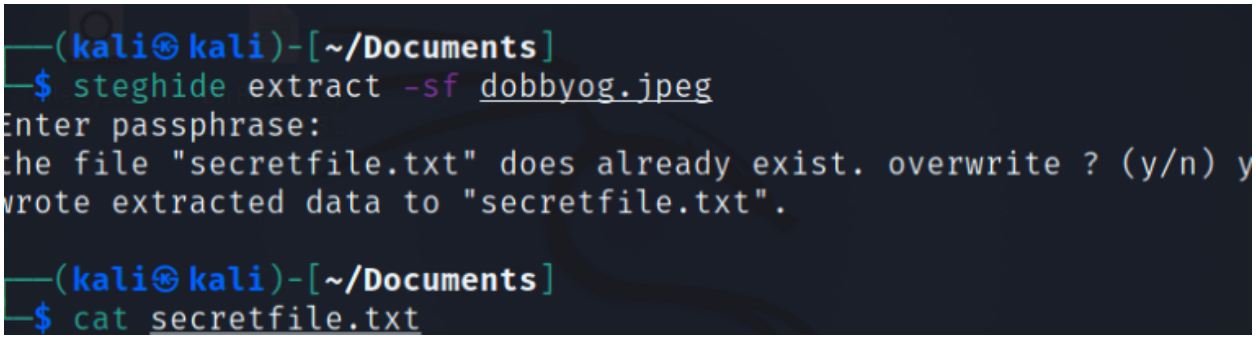


Figure 9.
And here is our secret file.

```
THIS IS A TOTALLY SECRET MESSAGE.

THIS IS FOR FULLSTACK ONLY.

AGAIN, THIS IS LIKE, TOTALLY CONFIDENTIAL:


HAPPY GRADUATION FRIENDS.
```

The reason that steganography is such a useful tool for malicious actors is because there is no sure fire way to defend against it or detect it. Some basic things to look for include overly large files, uneven bit mapping, or whether a bitmap image has a large number of duplicate colors. This would indicate that data has been embedded in the image. Looking at the file size and file properties will also tell you a lot. Anything unusual should raise immediate suspicion.

One way to try and detect the presence of steganography is simply looking for unusual patterns in a stego-image that seem to be obviously altered and suspicious. For example, examining unused areas on a disk or slack space. Lastly, if you are lucky enough to know what tool was used, you can obtain the same tool to compare the files. Or, rarely, if you have the original source file, you can do a comparison analysis. Remember that the art of steganography will be used where companies and/or governments will not allow one to encrypt

communications.
(https://www.giac.org/paper/gsec/2474/information-hiding-art-steganography/104295)

Can steganography transmissions be stopped on your network?

There are nefarious actors in the world that can and will attempt to send hidden information to malicious software also referred to as "stego-malware". This is of particular concern because open source, easy to use steganography tools are widely available, but tools to catch and prevent transmission are not. Examples of this are: Pingback, which uses the Internet Control Message Protocol (ICMP) sequence number for hidden communication, Convert-TCP, that modifies IP identification numbers and TCP initial sequence numbers (ISN), or the TCP/IP network steganography framework which embeds information into the ICMP identifier, ICMP sequence, IP identification field, and IP Differentiated services fields.
(https://dl.acm.org/doi/fullHtml/10.1145/3487405.3487421)

Filters are also available to apply to capture TCP/IP packets that contain hidden or invalid information in the packet headers. TCP/IP packets have unused space in the headers. The TCP header has six reserved or unused bits, and the IP packet header has two reserved bits. Information can also be hidden in the unused bits in the type of service (TOS) fields and flags of IP headers. This technique can be unsafe though because TCP/IP headers can get overwritten in the routing process and reserved bits could also be overwritten making the information useless.

The technology of firewalls has been improving. You can set filters to determine if packets are coming from within the firewall's domain. Also, with the validity of the SYN and ACK bits, the filters can be configured to catch packets that have information in presumed unused or reserved space, just like you can set certain firewalls to exclude such packets with spoofed addresses.

The work on steganalysis is vast and ongoing. As technology grows so do the ways that bad actors can grow with it. As augmented/virtual reality products expand in use, more than human deception is going to be a concern. Machines using AI and other technologies such as autonomously operated vehicles could be attacked in simply but effectively hidden malware attacks. Ransomware gangs and the growing cryptocurrency industry are also big areas of concern. Mal-adware can also somewhat easily be implemented using steganography. There are pros and cons to this issue. If you have malicious intentions it is relatively simple to get past security systems undetected. However, it is also difficult to transmit anything more than smaller amounts of information or data. If you are a company looking to protect your assets, steganography can be combined with cryptography to provide extra security for network communications and transmissions of information. This is a field that will continue to be part of

the security conversation because as security professionals solve a problem a bad actor will always be out there to create another!